

Turing Test for the Internet of Things

Neil Rubens^{1,2,3}

¹ Active Intelligence Lab

² Skolkovo Institute of Science and Technology

³ University of Electro-Communications

Abstract—Draft: Work in Progress:

How smart is your kettle? How smart are things in your kitchen, your house, your neighborhood, on the internet? With the advent of Internet of Things, and the move of making devices ‘smart’ by utilizing AI, a natural question arises, how can we evaluate the progress. The standard way of evaluating AI is through the Turing Test. While Turing Test was designed for AI; the device that it was tailored to was a computer. Applying the test to variety of devices that constitute Internet of Things poses a number of challenges which could be addressed through a number of adaptations.

I. INTRODUCTION

With the advent of the Internet of Things (IOT), and the move of making devices “smart” by utilizing AI [4], a natural question arises: “How can we evaluate the progress of making devices smarter?”. A common approach of evaluating AI is through the Turing Test (TT) [6] in which interrogator through a text conversation with two participants tries to determine which is a computer and which is a human (figure 1). However, in the context of the Internet of Things, Turing test (in its traditional form) may not be directly applicable in part due to the following:

- 1) Turing Test assumes that AI is embodied in a **single computer**:

“The present interest in ‘thinking machines’ has been aroused by a particular kind of machine, usually called an ‘electronic computer’ or ‘digital computer’. Following this suggestion we only permit digital computers to take part in our game.” [6]

- Internet of Things
 - multiple (connected) devices
 - many of the devices contain “computers”, but are not computers per-se (e.g. kettle or fridge).

- 2) Turing Test focuses on measuring **conversational intelligence**:

“The ideal arrangement is to have a teleprinter communicating between the two rooms.” [6]

- Internet of Things
 - some of the devices have no means of conducting a conversation (e.g. a toaster).

Evaluating AI in the context of IOT seems to require some adaptations, descriptions of which we provide in this paper. Similarly to the original Turing test paper [6]; we avoid defining “intelligence” and instead concentrate on practical/pragmatic aspects.



Figure 1. Turing Test: interrogator through a text conversation with two participants tries to determine which is a computer and which is a human [1].



Figure 2. Electric Kettle.

II. SIMPLIFIED TURING TEST FOR THE INTERNET OF THINGS

Let us propose a straightforward adaption of the TT to IOT; followed by describing the challenges that the test would face (in section §IV we propose extensions to TT that allow to alleviate at least some of the challenges). In this version we strive to make as few modifications to the original turing test as possible. Hence, we only modify the medium through which the “conversation” happens by replacing computer terminal with a “thing” (e.g. an electric kettle (figure 2)). To add the i/o capabilities, we add a remote input/control to the kettle (that controls the on/off switch) (similar to the keyboard in the

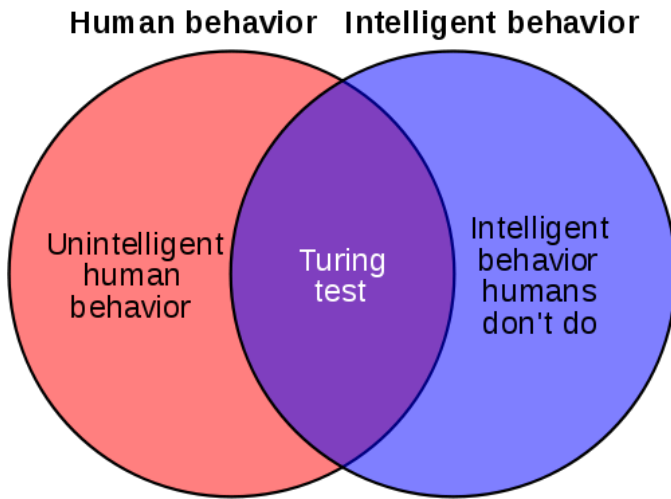


Figure 3. Turing Test: Intelligent Behavior vs Human Behavior [7].

original test). Interrogator is also able to interact with “kettle” through “local” on/off switch.

the communication channel (sending on/off status along with timestamps). One of the kettles would be controlled by a person and the other one would be controlled by a computer; interrogator’s task would be to decide which is which.

III. CHALLENGES FOR TURNING TEST IN THE CONTEXT OF THE INTERNET OF THINGS

Turing test has been a truly visionary work that became a standard for evaluating performance of AI systems; preceding by 6 years the official establishment of the field of Artificial Intelligence [2]. However, in 1950 it was difficult to foresee all of upcoming developments which would affect the development of AI. In this section, we describe various aspects of the Internet of Things that make it difficult to apply TT in a straightforward manner.

A. Limited Interface

In Turing test interaction is assumed to happen in the text conversation format. Many of the things constituting IOT have limited interfaces; e.g. a kettle has only on/off switch.

B. Limited Computational Capabilities

Many of the IOT devices have limited computing capabilities which are often geared to very specific purposes (e.g. auto-shutoff controller of the kettle).

C. Intelligent Behavior that is not Human Behavior

The Turing test is a test of a machine’s ability to exhibit intelligent behavior indistinguishable from that of a human. However, one of the major drawbacks of Turing Test is that not all of the intelligent behavior is human-like (figure 3). This is exacerbated if we consider that the test medium is not a textual conversation (which many things are not capable of functionally section §III-A). Below we outline some of the challenges that are exacerbated in the context of IOT.

1) *Analysis of Large Quantities of Numerical Data*: Most people may have difficulty with analyzing large quantities of numerical data (especially without analytical tools); more so if it comes from multiple sources.

2) *Response Time*: People are not able to respond as quickly as machines.

3) *Time: Availability / Duration*: Turing test assumes a relatively short pre-scheduled session during which examiner performs a test: “making the right identification after five minutes of questioning” [6]. The natural test duration for testing of things could be much longer; e.g. examining if your coffee making machine is intelligent could take weeks or months; and not necessarily at pre-scheduled times (e.g. if you happen to wake earlier than usual and need a sip of coffee).

4) *Memory / History*: In the original description it wasn’t stated explicitly if there is a conversational memory (script).

and we also add the conversation “memory” function that shows history of on/off activity (details ...); similar to the text conversation history displayed on the screen.

for this settings peoples memory is more limited than that of a computer

D. Evaluation Criteria

Turing test evaluates conversational intelligence. However it is not clear whether this criteria makes sense in different context; e.g. is it appropriate to measure how “intelligent” a kettle is by having a conversation with it; what if it talks well, but doesn’t perform well as a kettle (starts pondering the meaning of life, while deciding whether to boil the water)?

E. Multiple Participants

Turing test assumes interaction with a single participant at a time (be it a human or a computer). Internet of Things as the name implies, assumes that there are multiple entities involved. Therefore the challenge is adopting Turing test to multiple participants.

IV. PROPOSED APPROACHES

In this section we describe approaches to addressing challenges of adopting the Turing Test to the context of the Internet of Things (III).

A. Limited Interface

As described in (section §III-A) many of the things constituting IOT have limited interfaces; e.g. a kettle has only on/off switch. One might suggest that the original interface (e.g. on/off switch) could be used to do a binary encoding of a text conversation. Another option is to add an interface for a text conversation. However, even once interface allows for the textual conversation; other related challenges still remain unanswered: would the device computationally capable of having an intelligent conversation (section §III-B), and is conversation the proper way to evaluate the intelligence of a device (section §III-D). Answer to these questions might dictate the form of the interface. Our suggestion is to leave the interface as is.

I: day1: 8:00am : turn on kettle
 T: day1: 8:02am : turn off kettle
 I: day1: 8:10am : turn on kettle
 T: day1: 8:12am : turn off kettle

 T: day2: 7:58am : turn on kettle
 T: day2: 8:00am : turn off kettle
 I: day2: 8:10am : turn on kettle
 T: day2: 8:12am : turn off kettle

 T: day3: 8:09am : turn on kettle
 T: day3: 8:12am : turn off kettle

Figure 4. Interaction Intelligence Criteria. I - interrogator, T - participating thing (controlled by computer or a human) see section §IV-D1.

B. Limited Computational Capabilities

As described in (section §III-B) many of the devices have limited computational power and/or target for specific purposes (e.g. micro-controller in kettle for the auto-shut-off function). It is possible to modify/add hardware and software as to allow for generating and processing of textual conversations (either locally or remotely). However in this case, we might be appraising the intelligence of the add-ons rather than that of the device (which is our original goal). Hence our suggestion is to leave the hardware and software as is.

C. Intelligent Behavior that is not Human Behavior

While previous challenges were favorable to humans; the challenges in this section favors computers; since we focus on the intelligent behavior that humans don't do (section §III-C). We do want to make this a fair test hence we try to address these challenges. In the original paper [6] this challenge was pointed out:

“It is claimed that the interrogator could distinguish the machine from the man simply by setting them a number of problems in arithmetic. The machine would be unmasked because of its deadly accuracy.”

In the same paper [6] a potential solution was suggested:

“The machine (programmed for playing the game) would not attempt to give the right answers to the arithmetic problems. It would deliberately introduce mistakes in a manner calculated to confuse the interrogator.”

Making AI appear dumber and slower than it really is, has been used in programs aimed at passing the turing test, as well as in industry e.g. gaming [5].

“Dumbing down” approach could certainly be introduced in the context of IOT in a fairly straightforward manner. However, we strongly feel that the aim should be to achieve a high intelligence, rather than to simply mimic human behavior (section §IV-D). Hence we take a different approach of attempting to make human participants to appear more “computationally intelligent”. For the original test this could be achieved by equipping participants with calculators. In the context of IOT, additional modifications may be needed. To

narrow the gap between human and computers (for the tasks in which computers perform better); we consider two non-exclusive approaches: human-centered and computer assisted.

1) *Human-centered*: A typical approach of improving processing speed of computers is through:

- faster processing
- parallelization

People's response time could be improved in a similar manner:

- faster processing
 - using more skilled human operators
 - training human operators
- parallelization
 - using multiple participants

This approach does narrow the performance gap but not sufficient, and it is “expensive”. If this approach is used, it should also be complemented with the computer-assisted approach.

2) *Computer-assisted*: Receiving assistance from computers should allow to narrow (if not exceed) some of the performance gaps. In particular computers could be used to assist with: (1) analysis, (2) execution).

a) *Computer-assisted Analysis*: Using analytic software would allow to address the challenge of analyzing large quantities of numerical data (section §III-C1).

b) *Computer-assisted Execution*: The rest of the challenges (section §III-C) could be addressed by programming computers with the desired responses or analysis algorithms. In order to keep the distinction between pure AI and computer-assisted approach people should have ability to override computer's decisions.

D. Evaluation Criteria

Turing test evaluates conversational intelligence. However it is not clear whether this criteria makes sense in different context; e.g. is it appropriate to measure how “intelligent” a kettle is by having a conversation with it; what if it talks well, but doesn't perform well as a kettle (starts pondering the meaning of life for an hour, while deciding whether to boil the water)? On the other hand, focusing on evaluation operational performance does not necessarily reflect how “intelligent” the

thing is. Below we propose some of the approaches that could be used (in combination or independently).

1) *Interaction Intelligence* : In order to overcome some of the operational speed capabilities of humans and computers (section §III-C); we can evaluate intelligence through a form of dialogue. However, the dialogue is not be textual but is operational (figure 4). Both the interrogator and participants can control the time-stamps as to simulate the flow of time). Note that interrogator can prepare multiple dialogue scripts in order to examine the responses of the participants. In the end; based on the responses interrogator would have to decide whether the participant (interrogated thing) is controlled by a human or a computer.

2) *Operational Intelligence*: Devices could be evaluated based on how “intelligently” they behave. Unlike pure operational evaluation (e.g. how quickly the water is boiled); this criteria would also consider intelligence aspects, e.g. is a kettle able to distinguish between different users (members of the family), different usages (boiling water for tea, or for drip coffee, or other purposes).

3) *Inner Intelligence*: While the operational intelligence reflects the intelligence of the behavior of the objects (section §IV-D2); inner intelligence evaluation allows users to examine the internal rules/workings of the things to determine if those are intelligent or not. However this evaluation assumes that computer-assisted execution method was used (section §IV-C2), and that resulting rules could be represented in a human readable form e.g. if then rules or decision trees; however other methods are very hard to either create or examine manually e.g. SVM, random forests, neural networks.

E. Multiple Participants

Turning test assumes interaction with a single participant at a time (be it a human or a computer). Internet of Things, as the name implies, assumes that there are multiple entities involved. Therefore the challenge is adopting Turing test to multiple participants.

Assuming multiple participants raises more questions about evaluating:

- collective/distributed intelligence
- mixed groups
 - different devices
 - different types of entities (humans & non-humans)
- multiple evaluators
- many-many or one-many (e.g. one-many: one person controlling multiple things)
- participating vs non-participating entities (in the evaluation)

Naive approach would be to simply apply the single-version of the test to multiple objects and aggregate the results. More interesting approaches could be obtained by being somewhat creative; e.g. mixing human-controlled things with computer-controlled, etc.

V. CONCLUSION

It’s been over 50 years since the Turing Test for evaluating AI has been proposed. TT has stood up very well to the

test of times; and it is still by far the most prevalent way in which AI is evaluated. However, new settings, in particular the Internet of Things, pose challenges to the traditional definition of the Turing Test (section §III). We show that with some modifications (section §IV) the Turing Test could be adapted to these novel settings.

ACKNOWLEDGEMENT

This research is performed in conjunction with the Thing-Tank project [3] a part of the Ideas Lab MIT-Skoltech Initiative.

REFERENCES

- [1] J. Copeland. *Artificial Intelligence: A Philosophical Introduction*. Blackwell Publishers, Inc., Cambridge, MA, USA, 1st edition, 1993.
- [2] D. Crevier. *AI: The Tumultuous History of the Search for Artificial Intelligence*. Basic Books, 1994.
- [3] E. Giaccardi, C. Speed, and N. Rubens. Things making things. In *International Research Network for Design Anthropology*, 2014.
- [4] G. Kortuem and F. Kawsar. Smart objects as building blocks for the internet of things. *Internet Computing*, ..., 14(1):44–51, Jan. 2010.
- [5] L. Lidén. Artificial stupidity: The art of intentional mistakes. *AI Game Programming Wisdom*, 2:41–48, 2003.
- [6] A. Turing. Computing machinery and intelligence. *Mind*, LIX(236), 1950.
- [7] Wikipedia. Turing test — wikipedia, the free encyclopedia, 2014. [Online; accessed 25-November-2014].